We thank the three anonymous referees for their thorough reviews. Our replies are outlined in **red**.

1. Referee#1

1.1. General Comments

The paper describes the use of grape harvest data to reconstruct atmospheric circulation during the Little Ice Age. The approach is interesting, as are the conclusions that blocking situations were more likely in summer. However, the reconstruction approach is very indirect and an end-to-end error assessment is not done or not possible. In see this paper more as a stimulation of the discussion how historical proxies can be used quantitatively in paleoclimatology, but I have little faith in the outcome. Reconstruction the frequency of synoptic types based on annually resolved data is obviously extremely difficult.

1.2. Specific comments

p. 3027, l. 13: "Breakpoints could be documented". But what was then done?

This was not done for this study and hence this sentence will be removed.

p. 3028, Eq. 1: As I understand it, there are many ways or trajectories to arrive at F* for a given harvest date. The "inversion" of the model gives some kind of a best fit, but how good is it? It might be instructive to use some sort of Monte Carlo approach – using weather generators or even just using some climate model control run data to actually check how these different trajectories then in the end compare in terms of seasonal means. And of course it would be even more interesting to see how they compare in terms of blocking frequency.

There is only one way to obtain the harvest value. The inversion method uses the mean daily temperature of the reference period in each region and calculates the anomaly value for each year. The methodology does not to look for the daily value of the temperature, in which case it is necessary to use Monte-Carlo, or weather generator methods.

Thus, the quality and the uncertainty of the anomaly calculated depend of those already measured for the model (RMSE values shown in Table 2, in Supplementary material), which depend of the vineyard and the variety (the values vary between 6.1 and 10.5). The model quality obtained in this study is similar or better compared to errors obtained in other works about grapevine phenology (Williams et al 1985; Moncur et al 1989; Riou 1994; Oliveira

1998; (Jones 2003); van Leeuwen et al 2008; García de Cortázar-Atauri et al 2009; Caffarra & Eccel 2010; Duchêne et al 2010; Nendel 2010, Parker et al., 2011).

Caffarra, A. and Eccel, E. (2010) Increasing robustness of phenological models for *Vitis vinifera* cv. Chardonnay. International Journal of Biometeorology 54, 255-267.

Duchene, E., Huard, F., Dumas, V., Schneider, C. and Merdinoglu, D., 2010. The challenge of adapting grapevine varieties to climate change. Climate Research, 41(3): 193-204.

Jones, G. V. (2003) Winegrape Phenology. In Phenology: An Integrative Environmental Science, 1st edition. Ed. M.D Schwartz. (Kluwer Press: Massachusetts) pp. 523-539.

Moncur, M.W., Rattigan, K., Mackenzie, D.H. and McIntyre, G.N., 1989. Base Temperatures for Budbreak and Leaf Appearance of Grapevines. American Journal of Enology and Viticulture, 40(1): 21-26.

Nendel, C. (2010) Grapevine bud break prediction for cool winter climates. International Journal of Biometeorology 54, 231-241.

Oliveira, M., 1998. Calculation of budbreak and flowering base temperatures for Vitis vinifera cv. Touriga Francesa in the Douro Region of Portugal. American Journal of Enology and Viticulture, 49(1): 74-78.

Parker, A.K., Garcia de Cortazar-Atauri, I., van Leeuwen, C. and Chuine, I., 2011. General phenological model to characterise the timing of flowering and veraison of Vitis vinifera L. Australian Journal of Grape and Wine Research, 17(2): 206-216.

Riou C. (1994) The effect of climate on grape ripening: application to the zoning of sugar content in the European community. (European Commission: Luxembourg) pp. 319.

Van Leeuwen, C. et al., 2008. Heat requirements for grapevine varieties are essential information to adapt plant material in a changing climate. La connaissance des besoins en chaleur des principaux cépages de vigne est une donnée essentielle pour adapter le matériel végétal dans un contexte de changement climatique. . Compte rendus de VIIe Congrès International des terroirs viticoles, Changins, Switzerland, pp. 222-227.

Williams, D.W. et al., 1985. Validation of a Model for the Growth and Development of the Thompson Seedless Grapevine .2. Phenology. American Journal of Enology and Viticulture, 36(4): 283-289.

p. 3028, Eq. 1: It would be good to show the sensitivity of the model to temperatures graphically. The authors are looking at blocking situations in summer, which often are accompanied by heatwaves. RF in Eq. 1 gets very small for heatwaves. Does this make the model more senstitive to heatwaves (viz. blockings)?

There is no blocking (or reference to the atmospheric circulation) in Eq. (1): vine phenology responds to temperature only in our model. Wang and Engel's model (1998) was used to simulate veraison stage. This model has four parameters: a minimum, an optimum and a maximum temperature (Tmin, Topt, Tmax) and a threshold of cumulated temperature actions (F*) (dimensionless). Its curvilinear structure allows for the consideration of effects of high temperatures on development slowdown (see Figure below, obtained for three vine types).

Cardinal temperatures Tmin and Tmax, were fixed at 0°C and 40°C, respectively (Jones 2003), and parameters Topt and F* were fitted using the phenological dataset for each variety (Table 2 in Supplementary material, and figure R1 below).

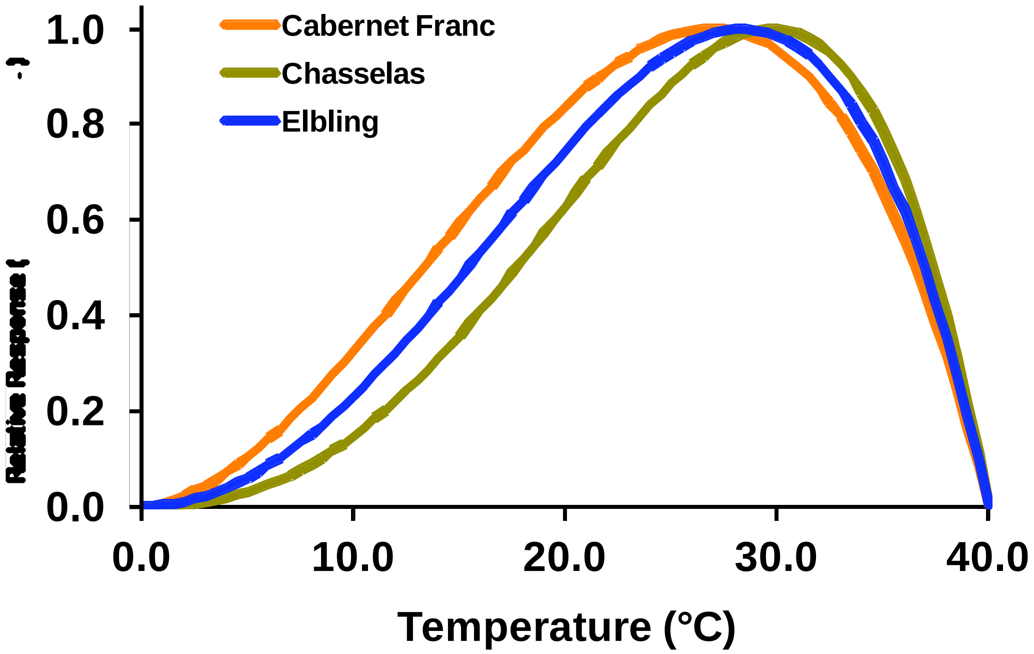

Figure R1: Temperature response (F*) as a function of temperature.

The use of the Wang and Engel (1988) model allows one to take into account high temperatures (i.e. heatwaves) that could not been observed using linear models. It allows for high temperatures beyond a threshold to slow down, rather than to hasten plant phenological development.

p. 3028, l. 8: Is t0 (15 March) assumed to be time independent; i.e., the same today than in the Little Ice Age. Can this be justified given all the work on spring phenology?

Grapevine is a perennial plant that needs cold temperatures to break dormancy release and to start a summation of daily heat requirements to reach the budbreak date. In literature, we find two different groups of models for grapevine: those that take into account chill requirements to simulate budbreak or other phenological stages (i.e. Garcia de Cortázar Atauri et al 2009; Caffarra & Eccel 2010); and those that consider that chilling requirements are very low, and start heat requirements computation from a fixed date (Williams et al 1985b; Nendel, 2010; Duchêne et al 2010; Parker et al., 2011) to simulate different phenological stages. First group of models are currently developed to study climate change impacts, it means, warmer winter conditions in the future. Second group of models were calibrated by different methods and in very different climates (usually in cold or very cold climates) obtained very similar starting dates (15 February; 1st March). In this context, and for our study about past climate, we have considered that t0 = 15 March (date obtained by calibration from our phenological database) has not been influenced, even during the Little Ice Age.

p. 3031, l. 2: How much do the correlations between different vineyards in the same region tell about the agreement of temperature reconstructions? These correlations only make sense to me if compared with the same correlations of the harvest dates.

Our sentence was not correctly formulated: we do not refer to inter-regional correlations, but correlations between reconstructed and observed temperatures within regions (and only mention the lower values of the correlations). This sentence will be corrected to clarify this point.

p. 3032, l. 8: I do not quite see what is done here. Percentiles of what? Does a quantification of the uncertainty of the gradient not require assumptions of the covariances of the errors?

We considered non-overlapping "windows" of 30 years and computed the 10^{th} and 90^{th} quantiles of temperature gradient reconstructions. For a centered Gaussian variable, this is equivalent to computing confidence intervals. As stated in the text, we determine a *confidence interval* around a climate reconstruction, similarly to what is done by (Luterbacher et al. 2004). This is not an estimate of *uncertainty*, strictly speaking. This will be clarified in the text.

p. 3032, l. 12: The correlations are quite low, they refer to the instrumental period (part of which is the calibration period), and they do not refer to the final product (atmospheric circulation).

This is what is stated in the text, but will be further emphasized.

p. 3034, l. 23: Why do you not directly compare your reconstruction with those of Küttel and Luterbacher?

This comparison will be provided (by adding a new panel in Fig. 4). A “draft” figure (figure R1) is provided below (the format will be improved for the resubmission). The mean SLP in 1750-1850 in the (Luterbacher et al. 2002) reconstruction yields a less pronounced anticyclonic pattern over western Europe, although SLP isolines are rather “flat” over France and western Europe.

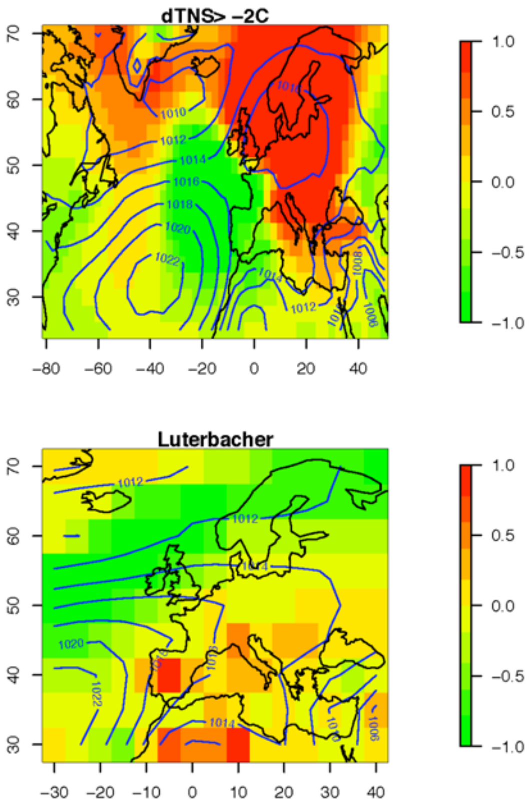

Figure R2: SLP patterns for steep and flat temperature gradients. The lower panel is the (Luterbacher et al. 2002) reconstruction between 1750 and 1850 (**note: the figure will look better in the resubmitted manuscript**).

p. 3034, l. 2: Maybe I am missing something here, but how do you come from the monthly mean gradients to a threshold for sampling individual days?

This is a methodological question that was not addressed in the manuscript, indeed. From Fig. 3 (or the 10 year averages of the N-S temperature gradient reconstructions), we have a heuristic idea of the range of persistent features of the temperature gradients. As Fig. 3 shows, the NS temperature gradients averaged over AMJJA, during the second half of the 20th century, are rather steep (~-3°C). Thus the AMJJA SLP in NCEP reanalysis and corresponding observed temperatures cannot sample the values that can be reached by the reconstructed temperature gradients prior to the 20th century. So, we used the daily data (between April and August), for which daily (observed) temperature gradients can be "flat" or "steep", in order to sample the whole range of possible values of the gradients and associated pressure patterns.

Similar as above, the computation of the circulation patterns could be done in a Monte Carlo-type approach to give an idea of the uncertainty. The gradients themselves have already no skill, but now you go one step further and do not quantify the errors anymore.

We prefer to use the distribution of the correlation coefficients between the patterns over which the composites are constructed (Fig. 5). This shows how the composite is representative of a typical situation conveyed by all patterns. A bootstrap-like approach would give a subset of the information contained in Fig. 5.

1.3. Technical comments

p. 3026, l. 11: Should this be "seven" rather than "eight"?

Corrected.

2. Referee#2

Yiou and co-authors present a study of continental-scale atmospheric circulation inferred from temperature recontructions using grape harvest dates from 4 regions in France and adjacent areas. The study is based on careful data collection and interpretation that have been carried out in multi-disciplinary project. Temperature reconstruction models based on inversed, process-based phenological models have been successfully reapplied to the newly available data.

2.1. General comments

Data and methods including calibration results and verification studies during the 20th century are clearly presented. However atmospheric circulation reconstructions from the Little Ice Age nor their comparisons with other reconstructions are very much hidden in the result part of the paper. I support the Editor's comment to include com-parisons of LIA circulation reconstructions with independent reconstructions. I even think that the comparisons should be shown in a figure.

We add a new figure with a comparison with the (Luterbacher et al. 2002) reconstructions (Fig. 4e, see reply to referee#1 above, figure R2).

Moreover I suggest that the temperature-atmospheric-circulation-relationship needs to be introduced and discussed more carefully throughout the manuscript. At the present stage I am not convinced that there is much more relevant information gained than 4 regional temperature reconstructions.

We make a more thorough explanation of the methodology in the revised manuscript. The basis of the atmospheric reconstruction is to use large scale "thermal" properties (rather that geostrophic properties). See Eq. (7.6b) in (Peixoto & Oort 1992).

The text suggest that the study provides new insights in circulation patterns at the continental European scale. However from Fig. 2 and 3 I am not convinced how reliable the results are. First, uncertainty estimates are missing.

We add a new panel to figure 1 (see below), in order to show the uncertainty in temperature reconstructions for each region. This uncertainty takes into account the intra-region variability and the number of available series within each region. The estimated standard deviation is weighed by the square root of the number of observations, as the standard deviation of the mean is proportional to the sample size.

Second, the changing number of GHD observations and its reduction back in time are not displayed in the reconstruction.

Plotting confidence intervals over figure 2 would make it extremely unreadable. Instead, we propose to include the uncertainty linked to the regional variance and the available number of series, as a complement to figure 1.

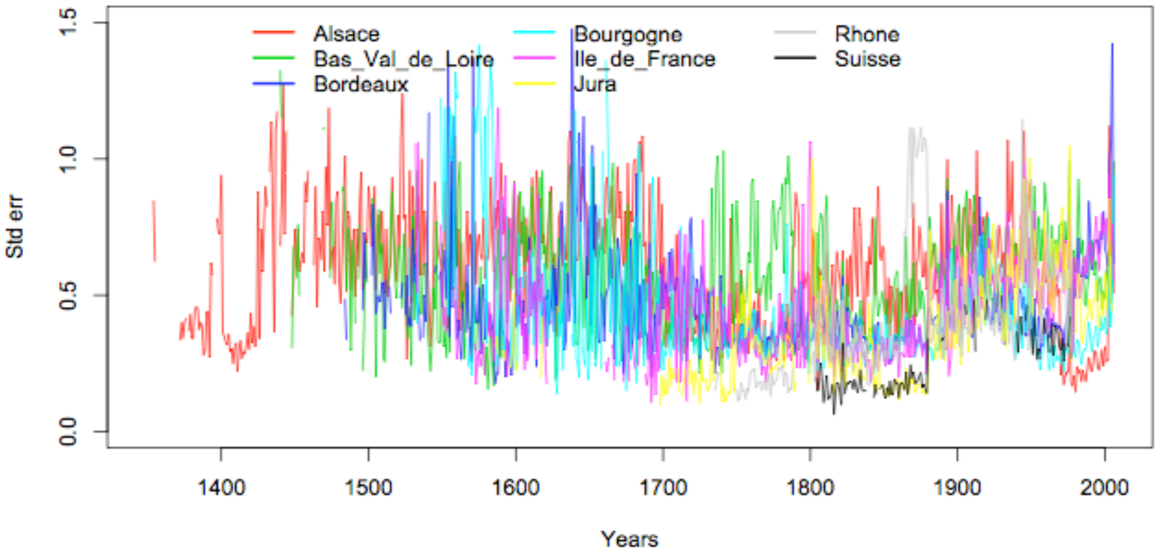

Figure R3: Amplitude of confidence intervals (1 sigma) of the reconstruction as a function of the number of available observations and their variability. The vertical axis is in °C (note: the final figure will look better).

Third, only the Eastern series has complete coverage during the LIA whereas West and North contain significant gaps.

Indeed, but our reconstruction procedure does not require continuous records, because it is a composite approach. We simply isolate decades when the NS temperature gradient is relatively flat. Time continuity is not an issue here.

In consequence the results in Fig 3 are overly positive.

We are not sure what the reviewer means. The confidence intervals do show how the NS temperature gradients vary. WE gradient variations are barely significant and do not change the features of the corresponding atmospheric pattern when the NS gradient is already > -2°C.

In the present form the study appears to be unfinished and lacks the conclusions that are indicated in the title. After completion of the work and a much clearer focus on circulation patterns the study will very much merit publication in Climate of the Past.

2.2. Minor comments

Introduction: maybe include paragraph p3033, l. 5ff here with more detailed descriptions.

P3026, 1st paragraph: what's exactly the benefot of so many co-authors?

This paper gathered competences from climatologists (PY, V. Daux, N. Viovy), ecophysiologists (I. Chuine, I. Garcia de Cortazar-Atauri), a historian (E. Garnier), and specialists in viticulture (C. van Leeuwen, A. Parker and JM Boursiquot).

P3027, l. 9ff: two personal communications are confusing. Did you use real ECAD download data? Method section: a clear description of the method for gradient constructions and inferences of pre-instrumental period atmospheric flow is missing.

We did use ECA&D data for Geneva, and Météo-France data for French stations because either they were not in ECA&D or they were not homogenized in ECA&D (e.g., Paris).

P3031: what is the importance of the second paragraph and data description for atmospheric circulation flow? Move to discussion or remove.

The paragraph states that the two southern regions yield comparable behavior in terms of grape harvest date trends. This does not seem to be an anthropogenic artifact, because wine grower practices did not changes (as shown in ancient agronomy documents) and the wine growers of the two regions did not have reasons to copy each other. This is a preamble to the use of those historical data to evaluate long terms changes.

3. Referee#3

Lines 11-13, p. 3028. The use of this uncertainty information in the reconstructions and analysis should be described here.
These values were used to determine the accuracy of the model and to calculate the uncertainties related to the anomaly reconstruction.

Line 18, p. 3028. Is this assumption of constancy reasonable?
This was discussed in Garcia de Cortazar- Atauri et al. (2011). This delay is not constant, but its standard deviation is 3 days at most.

Line 22, p. 3029. The use of these historical sources for information on the technical practices is quite valuable. The use of this uncertainty in the reconstructions and analysis should be described here.
To evaluate the uncertainty generated by past technical practices and non-thermal climate variables in each region, a range of inputs was provided to the STICS crop model: regional soil characteristics (field capacity, wilting point and bulk density), technical parameters (plant density, canopy geometry, fruit load, nitrogen fertilization, trimming date), variety information and weather daily data (maximum and minimum temperatures, rainfall, radiation, wind and humidity) as it was done in Garcia de Cortazar-Atauri et al. (2011).

Re: Fig. 2. Since these anomalies are constructed as a mean of the different varieties of the region, the standard deviations associated with this calculation should be shown in Figure 2.

The standard deviation will be shown in an additional panel in Figure 1 (see figure R3 above, in reply to referee 2).

Line 5, p. 3031. What is this consistency? It is not clear what this means, and more explanation needs to be provided.

We meant that the decadal trends are similar as in those reconstructions. This will be clarified in the revised manuscript.

Line 23, p. 3031. Why is this "certainly" the case? The argument provided appears to be more a supposition than evidence. This statement needs more thorough-going support. Also, it is not clear from Fig. 1 that there is a low number of GHD series in Southern France throughout this time period.

This is a hypothesis we formulate, because we do not have other explanations at hand in the old agronomy treaties. There is evidence that the phylloxera episodes were more severe in the south of France than in other regions.

Line 8, p. 3032. Are these percentiles taken from the 10th and 90th percentiles for T-opt and F* in Table 2? If not, then they need to be described better.

No, the percentiles are taken from the moving 30-year windows. This will be clarified in the manuscript.

ADD "approximately" before "correctly reproduced" in line 15, p. 3032.The statement as written suggests a level of accuracy that is not entirely supported by the results.

OK

REPLACE "minimal" with something like "at its reconstructed minimum" in line 25, p. 3032. The use of "minimal" here leads to incorrect interpretation. Its standard English interpretation in this context would indicate that the gradient is currently weak, when in fact it is reconstructed as strong.

OK

The statement in lines 16-18, p. 3034, is incorrectly stated, if one can interpret the anomaly information in a straightforward way. If this is the case, then this statement should read something more like the following:

"The EMULATE dataset shows a hint of a relative anticyclonic pattern over Western Europe, but much weaker and spatially far less extensive than the one inferred from the reconstructions in Fig. 4b-c. The associated low pressure anomaly structure immediately west

of Western Europe is much more like that inferred from the reconstructions, suggesting the existence of a blocking situation in terms of the relative E-W pressure gradient."
OK

Explain the consistency mentioned in lines 22-24, p. 3034, in more detail.
We add a new panel in Figure 4, showing the reconstruction of Luterbacher et al.

A brief paragraph should be added before the final paragraph in section 3.3, which would tie the circulation analogs estimated in this section to specific time periods for the 30-year average N-S and E-W gradient reconstructions that are discussed in section 3.2 (cf. Fig. 3).
We will add a paragraph that states that our reconstruction of SLP during a period of “shallow” NS temperature gradient is valid for ~1750-1850 AD.

REPLACE "in" before "wine" with "of", line 9, p. 3035.
OK

ADD "the" before "multi-decadal period" in line 19, p. 3035. MAKE "multi-decadal period" be plural in line 19, p. 3035.
OK

4. References

Jones GV (2003) Winegrape Phenology. In: Schwartz MD (ed) Phenology: An Integrative Environmental Science, Vol 39. Kluwer Press, Dordrecht p523 – 539

Luterbacher J, Dietrich D, Xoplaki E, Grosjean M, Wanner H (2004) European seasonal and annual temperature variability, trends, and extremes since 1500. SCIENCE 303:1499-1503

Luterbacher J, Xoplaki E, Dietrich D, Rickli R, Jacobeit J, Beck C, Gyalistras D, Schmutz C, Wanner H (2002) Reconstruction of sea level pressure fields over the Eastern North Atlantic and Europe back to 1500. Clim. Dyn. 18:545-561

Peixoto JP, Oort AH (1992) Physics of climate, Vol. American Institute of Physics, New York